Lehrstuhl für Regelungstechnik

Technical Reports on Automatic Control
Vol. TRAC-4, Nov. 2009
Available online at: www.rt.mw.tum.de

TECHNISCHE UNIVERSITÄT MÜNCHEN

How to choose a single expansion point in Krylov-based model reduction?

Rudy Eid*, Heiko Panzer, Boris Lohmann
Institute of Automatic Control, Technische Universität München
Boltzmannstr. 15, 85748 Garching, Germany

Abstract

In this short note, the problem of the choice of the expansion point in Krylov-based model reduction is addressed and the RK-ICOP algorithm for the automatic choice of a single expansion point for moment matching is revisited. This algorithm is numerically efficient, thus suitable for large-scale systems, and does not require any simulation or expert knowledge of the original system. In addition, a compact Matlab implementation is suggested.

1 Introduction

One of the most important parameters in Krylov-subspace methods is the expansion point about which the moments are matched. The value (and number) of the expansion point(s) steer the quality and the frequency range of the approximation of the transfer function, and the numerical effort needed to calculate the reduced order model.

Let s_0 be the point about which the Taylor series of the transfer function $H(s)$ of the system is expanded:

$$H(s) = -m_0 - m_1(s - s_0) - \cdots - m_i(s - s_0)^i - \cdots . \tag{1}$$

Then, the coefficients m_i are called moments [1, 3]. Accordingly, reducing a dynamical model while matching a number of moments (and/or the Markov parameters) about this point can be directly interpreted from a system theoretic point of view and employed to describe the similarity between the original and reduced models based on the following facts:

1. With $s_0 = 0$, the reduced and original model have the same DC gain, and steady state accuracy is achieved.
2. Small values of s_0 result in a reduced model with a good approximation of the slow dynamics.
3. Large values of s_0 (and/or matching the Markov parameters) result in a reduced model approximating the system's behavior at high frequencies.
4. When matching some of the moments about different frequency points simultaneously, a better approximation on a wider frequency band or on a specific frequency band of interest can be achieved.

*Corresponding author: Tel: +49 89 289 15592, Fax: +49 89 25915659, email: eid@tum.de

Even though these facts give an idea on the choice of the expansion point, no specific value of s_0 can be derived based on them.

Consequently, the choice of the interpolation point(s) in moment matching became and is still an active field of research. In [3], the basic principles for this choice have been established and the different possibilities involving real and/or imaginary, single and multiple expansion points, have been considered and discussed:

1. A purely imaginary expansion point leads to very good local approximation and to a very slow convergence at all frequencies away from s_0.
2. A real interpolation point offers good approximation in a large neighborhood around s_0, except around some lightly damped eigenvalues on the imaginary axis.
3. The combination of several real, imaginary, and/or complex expansion points can guarantee a good approximation of the transfer function in multiple frequency regions.

In the literature, different possibilities for the choice of the single or multiple interpolation points have been presented, targeting different aims. In [9], the problem of passivity preserving order reduction has been addressed and a rational Krylov algorithm with q interpolation points selected as spectral zeros of the original transfer function has been presented. In [4], an iteratively corrected rational Krylov algorithm (IRKA) for H_2 model reduction of SISO systems has been suggested. This interpolation-based idea consists of matching only one moment about each of the q expansion points chosen at the mirror image of the poles of the reduced system. These results have been generalized to the MIMO and $H_{2\alpha}$ cases. For more details, refer to [6] and the references therein. The determination of appropriate starting values and the convergence of this iterative algorithm are still considered as open questions. In addition, the complexity and numerical costs of these approaches are significant, when compared to a reduction with a single expansion point and a relatively high reduced order. Moreover, the methods above do not consider the time responses in any form but only the frequency response.

Therefore in practice, the choice of s_0 is still often an ad-hoc process, where the user has to try several expansion points (or several ones simultaneously) and pick up the one(s) resulting in a satisfactory approximation for his system. The choice $s_0 = 0$ is widely used and comes generally at the first place as it often delivers good results in a large neighborhood of the low-frequency part of the spectrum, including the steady state.

In this note, we propose a simple and numerically efficient procedure for calculating a suitable value for the single expansion point s_0. The optimality properties of this expansion point result from time-domain considerations of the system response and are explained in detail in [1]. Roughly speaking, using the the equivalence property of moment matching and Laguerre-based order reduction [2], the problem is reformulated as finding the best choice for the free parameter α in the Laguerre basis function. Minimizing an appropriate cost functions with very few iterations is the key point towards finding an optimal interpolation point s_0 that guarantees a good approximation of all the time responses of the system and the peak area of the Bode diagram [1].

2 An *optimal* expansion point for moment matching

Consider the SISO LTI dynamical system of the form:

$$\left\{ \begin{array}{rcl} \dot{\mathbf{x}}(t) & = & \mathbf{A}\mathbf{x}(t) + \mathbf{b}u(t), \\ y(t) & = & \mathbf{c}^T\mathbf{x}(t), \end{array} \right. \tag{2}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{c} \in \mathbb{R}^n$ are constant matrices, $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$ and $\mathbf{x}(t) \in \mathbb{R}^n$ are, respectively, the input, output and states of the system.

The aim of moment matching is to find a reduced order model of order $q \ll n$, whose moments match some of those of the original one [3, 10].

One way to calculate this reduced order model is applying a projection to the original model (2) (here, $\mathbf{W}^T\mathbf{V} = \mathbf{I}$ is assumed to simplify the presentation),

$$\left\{\begin{array}{rcl} \dot{\mathbf{x}}_r(t) &=& \overbrace{\mathbf{W}^T\mathbf{A}\mathbf{V}}^{\mathbf{A}_r}\mathbf{x}_r(t) + \overbrace{\mathbf{W}^T\mathbf{b}}^{\mathbf{b}_r}u(t), \\ y(t) &=& \underbrace{\mathbf{c}^T\mathbf{V}}_{\mathbf{c}_r^T}\mathbf{x}_r(t), \end{array}\right. \tag{3}$$

by means of the so-called projection matrices $\mathbf{V}$ and $\mathbf{W} \in \mathbb{R}^{n\times q}$. These projection matrices are chosen to span bases of the input and output Krylov subspaces,

$$\mathcal{K}_q\left((\mathbf{A}-s_0\mathbf{I})^{-1}, (\mathbf{A}-s_0\mathbf{I})^{-1}\mathbf{b}\right), \tag{4}$$

$$\mathcal{K}_q\left((\mathbf{A}-s_0\mathbf{I})^{-T}, (\mathbf{A}-s_0\mathbf{I})^{-T}\mathbf{c}\right). \tag{5}$$

which allows to match the first $2q$ moments around s_0[1]. For the numerical calculation of $\mathbf{V}$ and $\mathbf{W}$, the known Lanczos or Arnoldi algorithms or one of their improved versions can be used.

It was shown in [1] that one good choice for the expansion point s_0 is taking it equal to the optimal Laguerre parameter α that guarantees an optimal approximation of the impulse response $h(t)$ of the original system.

$$s_0 = \alpha = \sqrt{\frac{M_2}{M_1}}, \tag{6}$$

$$\text{with} \quad M_1 = \frac{\int_0^\infty t\,h^2(t)dt}{\int_0^\infty h^2(t)dt}, \quad \text{and} \quad M_2 = \frac{\int_0^\infty t\,\dot{h}^2(t)dt}{\int_0^\infty h^2(t)dt}. \tag{7}$$

Based on their definition, the parameters M_1 and M_2 can be physically interpreted as being, respectively, the *decay rate* and *smoothness* of the system's impulse response. For more details related to the optimization problem, the Laguerre-based framework and its connection to Krylov-based model reduction refer to [1].

The main question arising in this context is the calculation of the optimal parameter $s_0 = \alpha$ in practice. Although the problem looks complicated in the general case, it may be easily solved for LTI systems, leading to:

$$s_0 = \alpha = \sqrt{\frac{\mathbf{c}^T\mathbf{A}\mathbf{Y}\mathbf{A}^T\mathbf{c}}{\mathbf{c}^T\mathbf{Y}\mathbf{c}}}, \tag{8}$$

where $\mathbf{Y}$ is the solution of the Lyapunov equation,

$$\mathbf{A}\mathbf{Y} + \mathbf{Y}\mathbf{A}^T + \mathbf{X} = \mathbf{0}, \tag{9}$$

and $\mathbf{X}$ is the controllability gramian, i. e.

$$\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^T + \mathbf{b}\mathbf{b}^T = \mathbf{0}. \tag{10}$$

[1] In the so called *one-sided* method, only one Krylov subspace is used with a common choice $\mathbf{W} = \mathbf{V}$, and only q moments match.

Remark 1. If the impulse response data is already available, for instance, after a system simulation, the calculation of M_1 and M_2 is then numerically possible and straightforward. All what is needed is to numerically evaluate the integrals involved in the definition of these parameters (7) and the Lyapunov equations need not to be solved.

3 The RK-ICOP Algorithm

If the impulse response data is not available, the calculation of s_0 from (8) will be costly as two Lyapunov equations in the size of the original system are to be solved. By finding *approximate* solutions of the required Lyapunov equations, the cost of calculation can be dramatically reduced. This is achieved by approximately calculating $\mathbf{X}$ and $\mathbf{Y}$ using a reduced system by means of the so-called Galerkin conditions [5, 8].

Accordingly, the optimal parameter is then approximated as,

$$s_0 \approx \sqrt{\frac{\mathbf{c}^T\mathbf{A}\mathbf{V}\mathbf{Y}_r\mathbf{V}^T\mathbf{A}^T\mathbf{c}}{\mathbf{c}^T\mathbf{V}\mathbf{Y}_r\mathbf{V}^T\mathbf{c}}}, \tag{11}$$

or alternatively, as [7]:

$$s_0 \approx \sqrt{\frac{\mathbf{c}_r^T\mathbf{A}_r\mathbf{Y}_r\mathbf{A}_r^T\mathbf{c}_r}{\mathbf{c}_r^T\mathbf{Y}_r\mathbf{c}_r}}. \tag{12}$$

Such an approximation depends on the reduced system itself, and hence, the optimization is not straightforward. To converge to an optimal solution, it is proposed to iterate between the optimal parameter and the reduced system starting from an initial parameter. This results in the RK-ICOP algorithm:

Algorithm 1. Rational Krylov with an Iteratively Calculated Optimal Point (RK-ICOP)

1. Reduce the original system (2) to order q using an arbitrary value s_{0_0} and set $i = 1$.
2. For the reduced system (3), solve the Lyapunov equations:
$$\mathbf{A}_r\mathbf{X}_r + \mathbf{X}_r\mathbf{A}_r^T + \mathbf{b}_r\mathbf{b}_r^T = \mathbf{0}, \tag{13}$$
$$\mathbf{A}_r\mathbf{Y}_r + \mathbf{Y}_r\mathbf{A}_r^T + \mathbf{X}_r = \mathbf{0}, \tag{14}$$
3. Calculate the approximation of the optimal parameter s_{0_i} using (11) or (12).
4. Reduce the original system using s_{0_i}.
5. Increase i and go back to step 2.

The algorithm may be terminated if $\frac{|s_{0_i} - s_{0_{i-1}}|}{s_{0_i}} \leq \epsilon$ for a given tolerance ϵ. In most practical applications, not more than 3 iterations were needed.

Remark 2. Note that eq. (11) and (12) are valid for both, one- and two-sided reduction methods, without changes [7].

The key idea of this algorithm lies in the fact that it tries to minimize the effect of the higher order terms in the infinite Laguerre series expansions of the impulse response, and that it

simultaneously considers the frequency-based moment matching approach from a time-domain perspective. Moreover, the algorithm has a simple structure, is numerically efficient, and thus suitable for the reduction of large-scale systems (see next section).

Numerous simulations have shown that the initial value s_{0_0} almost does not affect neither the convergence in general nor its speed, nor the accuracy of the result. In addition, the RK-ICOP has the feature of calculating an expansion point that tends to detect the dominant peak in the Bode diagram without the need of calculating or giving any hint about this diagram [1, 7].

4 A Matlab Implementation

Below a commented and ready-to-run version of the ICOP algorithm for Matlab [1, 7]:

```
function [sysr V s0] = ICOP(sys, q, s0)

%RK-ICOP Algorithm
%for the determination of a single optimal expansion point in Krylov-based MOR
%
% [sysr V s0] = ICOP(sys, q, s0)
%
% Inputs:  * sys:  Large-Scale system SISO, as 'ss' variable
%          * q:    Order of the reduced system
%          * s0:   Initial expansion point (starting value for iteration)
%
% Outputs: * sysr: Reduced system (here using a one-sided Krylov method) using the
%                  optimal expansion point
%          * V:    Projection matrix spanning the Krylov subspace
%          * s0:   Optimal expansion point
% % ---- Begin example ----
% sys=ss(A,B,C,D);
% [sysr V s0] = ICOP(sys, q, 0);
% % ---- End example ----
% Authors: Rudy Eid, Heiko Panzer  {eid,panzer}@tum.de (www.rt.mw.tum.de)
% Date:  29 Oct 2009

% Get the system matrices
A = sys.A; B = sys.B; C = sys.C; D = sys.D;

for i=1:4  % max. number of iterations
      % 1. Calculate the projection matrix using the Arnoldi algorithm
           V=arnoldi(eye(size(A)), A-s0*eye(size(A)), B, q);
           % Calculate the reduced system via projection for the actual expansion point
           Ar=V'*A*V; Br=V'*B; Cr=C*V;
           sysr = ss(Ar,Br,Cr,D);

       % 2. Calculate the new expansion point
           % Solve the Lyapunov equations in the reduced dimension
           Xr=lyap(Ar,Br*Br');
           Yr=lyap(Ar,Xr);

           % calculate the new s0
           s0_old = s0;
           s0=sqrt((Cr*Ar*Yr*Ar'*Cr')/(Cr*Yr*Cr'));

       % 3. Check for convergence
           if ( norm(s0-s0_old)/norm(s0) ) < 1e-2
               break
           end
end
```

For more MATLAB codes and further information, please visit our website: www.rt.mw.tum.de

References

[1] R. Eid. *Time domain model reduction by moment matching.* PhD thesis, Institute of Automatic Control, Technische Universität München, 2009.

[2] R. Eid, B. Salimbahrami, and B. Lohmann. Equivalence of laguerre-based model order reduction and moment matching. *IEEE Trans. on Automatic Control*, 52(6):1104–1108, 2007.

[3] E. J. Grimme. *Krylov Projection Methods for Model Reduction.* PhD thesis, Dep. of Electrical Eng., Uni. Illinois at Urbana Champaign, 1997.

[4] S. Gugercin, A. C. Antoulas, and C. Beattie. H2 model reduction for large-scale dynamical systems. *SIAM Journal on Matrix Analysis and Applications,*, 30(2):609–638, 2008.

[5] I. M. Jaimoukha and E. M. Kasenally. Krylov Subspace Methods for Solving Large Lyapunov Equations. *SIAM J. Numer. Anal.*, 31(1):227–251, 1994.

[6] D. Kubalinska. *Optimal interpolation-based model reduction.* Ph.D thesis, University of Bremen, 2008.

[7] H. Panzer. Entwicklung eines vollautomatischen Krylov-basierten Verfahrens der Modellordnungsreduktion. Diplomarbeit, Institute of Automatic Control, Technische Universität München, Munich, Germany, 2009.

[8] Y. Saad. Numerical Solution of Large Lyapunov Equations. In Kaashoek, M., Schuppen, J. V., and Ran, A., editors, *Signal Processing, Scattering, Operator Theory and Numerical Methods*, pages 503–511, Boston, MA, 1990. Birkhauser.

[9] D. C. Sorensen. Passivity preserving model reduction via interpolation of spectral zeros. *Sys. and Cont. Letters*, 54:347–360, 2005.

[10] C. D. Villemagne and R. E. Skelton. Model Reduction using a Projection Formulation. *Int. J. Control*, 46:2141–2169, 1987.